

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:52:28 ; Search time 170.72 Seconds
(without alignments)
17.102 Million cell updates/sec

Title: US-09-331-631A-3_COPY_74_116
Perfect score: 250
Sequence: 1 NODDPQDCQQCQRRCKRQOE.....RQOQYCCRRCKEICEEEY 43

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_66:**
1: PIR1:**
2: PIR2:**
3: PIR3:**
4: PIR4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	112	44.8	566	2 S22477	vicilin precursor
2	102	40.8	509	2 S08059	alpha-globulin typ
3	100	40.0	588	1 FMCNAB	alpha-globulin B p
4	92	36.8	411	2 T29475	hypothetical prote
5	90	36.0	605	2 S06398	alpha-globulin typ
6	78	31.2	572	2 T29880	hypothetical prote
7	77.5	31.0	425	2 T18592	hypothetical prote
8	77.5	31.0	600	2 T18593	hypothetical prote
9	77.5	30.8	810	2 T44430	protein PVI100 (imp
10	75	30.0	47	2 JCS557	arginine/glutamate
11	75	30.0	154	2 T27967	hypothetical prote
12	73.5	29.4	242	2 T29699	hypothetical prote
13	73	29.2	111	2 T20452	hypothetical prote
14	72	28.8	330	2 T25169	hypothetical prote
15	71	28.4	388	2 T31887	hypothetical prote
16	71	28.4	388	2 T31888	hypothetical prote
17	71	28.4	438	2 T31889	hypothetical prote
18	71	28.4	445	2 T31898	hypothetical prote
19	69.5	27.8	335	2 T31560	hypothetical prote
20	69.5	27.8	335	2 T31561	hypothetical prote
21	67.5	27.0	314	2 T27686	hypothetical prote
22	67.5	27.0	335	2 T31559	hypothetical prote
23	67	26.8	1513	2 T23681	hypothetical prote
24	66.5	26.6	342	2 JC7110	brain-specific mem
25	66.5	26.6	637	2 S35221	hypothetical prote
26	66	26.4	165	2 T20468	hypothetical prote
27	64.5	25.8	654	2 T30136	hypothetical prote
28	63.5	25.4	33	2 A41822	antimicrobial pept
29	63.5	25.4	273	2 T20930	hypothetical prote

30	63.5	25.4	284	2 S23256	tropomyosin beta -
31	63.5	25.4	1306	2 T13592	hypothetical prote
32	62.5	25.0	336	2 S31574	hypothetical prote
33	62	24.8	419	2 T18450	hypothetical prote
34	61.5	24.6	51	1 HSM851	protamine - mouse
35	61.5	24.6	51	2 S03997	protamine 1 - rat
36	61	24.4	910	2 A34721	androgen receptor
37	61	24.4	911	2 B34721	androgen receptor
38	61	24.4	919	2 A39248	androgen receptor
39	59.5	23.8	294	2 T23682	hypothetical prote
40	59	23.6	644	2 A36325	epidermal growth f
41	59	23.6	1094	2 S49313	protein kinase - s
42	59	23.6	2318	2 S45306	notch 3 protein -
43	58.5	23.4	1282	2 JE0120	glycoprotein A - m
44	58	23.2	447	2 I38975	nuclear orphan rec
45	58	23.2	648	1 J01150	protein kinase (EC

ALIGNMENTS

RESULT 1

S22477
vicilin precursor - cacao
C:Species: Theobroma cacao (cacao)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999
C:Accession: S22477; S22478; S18105; S22050
R:McHenry, L.; Fritz, P.J.
Plant Mol. Biol. 18, 1173-1176, 1992
A:Title: Comparison of the structure and nucleotide sequence of vicilin genes of coco
A:Reference number: S22477; MUID:92288309
A:Accession: S22477
A:Molecule type: DNA
A:Residues: 1-566 <MC3>
A:Cross-references: EMBL:X62625
A:Accession: S22478
A:Molecule type: mRNA
A:Residues: 1-452 <MC2>
A:Cross-references: EMBL:X62626
C:Genetics:
A:Introns: 211/1; 269/3; 296/3; 391/3; 502/1
C:Superfamily: glycinin
C:Keywords: seed; storage protein
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-566/Product: vicilin #status predicted <MAT>

Query Match

Best Local Similarity 44.8%; Score 112; DB 2; Length 566;
Matches 19; Conservative 10; Mismatches 11; Indels 0; Gaps 0;

QY 2 ODDPQDCQQCQRRCKRQESGPRQOQYCCRRCKEICEEE 41
DB 78 EEELQROVQDCQGRCEQDQOQGRQOQCCRRCKWEYKDE 117

RESULT 2

S08059
alpha-globulin type B precursor (tandem 1) - upland cotton (fragment)
N:Alternate names: seed storage protein
C:Species: Gossypium hirsutum (upland cotton)
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
C:Accession: S08059
R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 9, 533-546, 1987
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX.
A:Reference number: S08059
A:Accession: S08059
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-509 <CHL>
C:Superfamily: glycinin

C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18592
R:McMurray, A.
Submitted to the EMBL Data Library, April 1996
A:Reference number: Z18995
A:Accession: T18592
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1425 <WIL>
A:Cross-references: EMBL:Z71177; PIDN:CMA94867.1; GSPDB:GND0023; CESP:AC3.3
A:Experimental source: clone AC3
C:Genetics:
A:Gene: CESP:AC3.3
A:Map position: 5
A:Introns: 18/3
C:Superfamily: g1ladin

	Query Match	31.0%;	Score 77.5;	DB 2,	Length 425;
	Best Local Similarity	40.0%;	Pred. No. 0.33;		
	Matches	18;	Conservative	9;	Mismatches 7; Indels 11; Gaps 4.
Oy	6 OTDCC-----OCORRCR---QDSGSRDOOYCGRCKE---ICEEE 40				
Db	236 OTTCQAVPQCQQCCAPCQCPSPSAVCCCO--CNTCCQAAPVCQQO 279				

RESULT 8
T18593
hypothetical protein AC3.4 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18593
R:McMurray, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z18995
A:Accession: T18593
A:Status: preliminary: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-600 <MIL>
A:Cross-references: EMBL:Z71177; PIDN:CAA94868.1; GSPDB:GND00023; CESP:AC3.4
C:Experimental source: clone AC3
C:Genetics:
A:Gene: CESP:AC3.4
A:Map position: 5
A:Introns: 292/2; 374/3; 572/2

Query Match	31.0%;	Score 77.5;	DB 2;	Length 600;
Best Local Similarity	40.0%;	Pred. No. 0.42;		
Matches 18; Conservative		9; Mismatches	7; Indels	11; Gaps 4

QY	6	OTDQ----	QCRRR--	QESGRRQOQY	QCRRRK----	ICEEE	40
Db	103	OTTCQAAVPCQCCQCAPQCCQSPASQCCQ--	COMTCQQAAPVCOQ				146

RESULT 9
T44430
protein PVI00 [imported] - winter squash
C:Species: Cucurbita maxima (winter squash)
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T44430
R:Yamada, K.; Shimada, T.; Kondo, M.; Nishimura, M.; Hara-Nishimura, I.
J. Biol. Chem. 274, 2563-2570, 1999
A:Title: MULTIPLE functional proteins are produced by cleaving Asn-Gln bonds of a single
A:Reference number: 722767; MUID:99107919
A:Accession: T44430
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-810 <YAM>
A:Cross-references: EMBL:AB019195; NID:q3808061; PTDN:BAA34056.1; PID:q3808062

	DB	Score	77;	DB	2;	Length	810;
	Best Local Similarity	34.2%;	Pred.	No.	0.55;		
Matches	13;	Conservative	10;	Mismatches	15;	Indels	0;
Gaps	0;						

```

RESULT 10
JC5557
  arginine/glutamate-rich 6.5K polypeptide - smooth loofah
C:Species: Luflia cylindrica (smooth loofah)
C:Date: 02-Sep-1997 #sequence-revision 05-Sep-1997 #text_update 07-May-1999
C:Accession: JC5557
R:Kikunaga, M., Park, S.S., Sakai, R., Yamasaki, N., Funatsu, G.
  Biosci. Biotechnol. Biochem. 61, 984-988, 1997
A:Title: Primary structure of 6.5K-arginine/glutamate-rich polypeptide from the seeds
A:Reference number: JC5557; MUID:97357433
A:Accession: JC5557
A:Molecule type: protein
A:Residues: 1-47 <KIM>
A:Experimental source: seed
A:Comment: This protein is a storage protein which provides nitrogen and carbon reser
F:12-33,16-29/Disulfide bonds: #stratus predicted

```

	Query Match	30.0%	Score 75;	DB 2;	Length 47;
	Best Local Similarity	37.1%	Pred. No. 0.13;		
	Matches	13; Conservative	10; Mismatches	8; Indels	4; Gaps
Oy	5	PQTGCGCQRCRCROESPPROOQCRCRKECEE	39		
	I:I:::	I::I:::	I::I:::		
Dd	5	FRIETACRCQAHEGV-----QRCCOOVCE	35		

RESULT 11
T27967
hypothetical protein ZK673.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27967
R:McMurray, A.
submitted to the EMBL Data Library, March 1995
A:Reference number: Z20447
A:Accession: T27967
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-154 <ML>
A:Cross-references: EMBL:Z48585; PIDN:CMA8486.1; GSPD:GN00020; CESP:ZK673.1
A:Experimental source: clone ZK673
C:Genetics:
A:Gene: CESP:ZK673.1
A:Map position: 2
A:Introns: 64/1; 93/2; 119/2

	30.0%;	Score 75;	DB 2;	Length 154;
Query Match:	Best Local Similarity	41.9%;	Pred. No. 0.28;	
Matches	13; Conservative	3; Mismatches	15; Indels	0; Gaps
QY	3 DDFQTDCCGCCRRCQGSGSPROOYCQRRC	33		
	69 DDNDTDCYTSLCSNAKTYPLDGFCKTC	99		

RESULT 12
T29699
hypothetical protein f31a3.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000

```

C:Accession: J129639
R:Murray, J.; Le, T.T.
submitted to the EMBL Data Library May 1996
A:Description: The sequence of C. elegans cosmid F31A3.
A:Reference number: Z20667
A:Accession: J29639
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-242 <MOR>
A:Cross-references: EMBL:U58742; PIDN:AAB36856.1; GSPDB:GN00028; CESP:F31A3.1
A:Experimental source: strain Bristol N2; clone F31A3
C:Genetics:
A:Gene: CESP:F31A3.1
A:Map position: X
A:Introns: 18/3; 160/3
C:Superfamily: ultra-high-sulfur keratin

Query Match      29.4%; Score 73.5; DB 2; Length 242;
Best Local Similarity 33.3%; Pred. No. 0.57;
Matches 17; Conservative 10; Mismatches 13; Indels 11; Gaps 4.

QY      2  ODDPDTDOCCOCORRCR-----QDSGPP-QOQY--CORRCKEICEEEE 42
      1  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DbB      77  QQPAPAPCCQCCNTCCQCSAPVCQQQCCAPCCQCCQFAPACQ-QCCNSCCQQTQ 126

```

```

RESULT 13
T20452
hypothetical protein E04D5.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T20452
R:McMurray, A.
submitted to the EMBL Data Library, October 1995
A:Reference number: Z19277
A:Accession: T20452
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-111 <WIL>
A:Cross-references: EMBL:Z66496, PIDN:CAA91281.1; GSPDB:GN00020; CESP:E04D5.4
A:Experimental source: clone E04D5
C:Genetics:
A:Gene: CESP:E04D5.4
A:Map position: 2
A:Introns: 56/1; 85/2

Query Match 29.2%; Score 73; DB 2; Length 111;
Best local Similarity 41.9%; Pred. NO. 0.36;
Matches 13; Conservative 3; Mismatches 15; Indels 0; Gaps 0.

0Y 3 DDPTDCCGCGRCRQGESGPRQOQYCCRC 33
||| ||| | | : : ||: | : |
DB 61 DDPTDCCGCTGYTFLCSNAKTYPLDQFCAKTC 91

```

RESULT 14
T25169
hypothetical protein T23FL.6 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T25169
R:Wilkinson, J.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19990
A:Accession: T25169
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Kolecule type: DNA
A:Residues: 1-330 <WIL>
A:Cross-references: EMBL:Z81129; PIDD:CAB03405.1; GSPDB:GN00023; CESP:T23FL.6
A:Experimental source: clone T23FL
C:Genetics:

A:Gene: CESEP:T23F1.6
A:Map position: 5
A:Introns: 16/3
C:Superfamily: gliadin

Query Match	28.8%	Score 72;	DB 2;	Length 330;
Best Local Similarity	29.5%;	Pred. NO. 1;		
Matches	13;	Conservative	10;	Mismatches 9;
				Indels 12;
				Gaps 2
Oy	7	TDDCCGCGRCRQ-----	ESGPRQD-----	YCARCKEICE 38
Db	230	TTCGCGCNSCNSNTOTITTYVQASPTQSCVPCGCGGCGCGCGC		273

```

RESULT      15
T31887
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C:Accession: T31887
R:Greco, T.; Bradshaw, H.; Elliott, G.
A:Submitted to the EMBL Data Library, July 1997
A:Description: The sequence of C. elegans cosmid C03A7.
A:Reference number: Z21096
A:Accession: T31887
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-388 <GRF>
A:Cross-references: EMBL:AF016451; PIDN:AAB66001.1; GSPDB:GN00023; CESP:C03A7.4
A:Experimental source: strain Bristol N2; clone C03A7
C:Genetics:
A:Gene: CESP:C03A7.4
A:Map position: 5
A:Introns: 75/3
C:Superfamily: giliadin

```

Q7	6	QTDCC-QCCARC-----RQESGPRQOQYCCARKCEICEE	39
Db	194	QAGCCPQCQCCQCCQSSCVDQDQSSNCCPCPCNCCGCSDDICQ	233
		Query Match 28.4%; Score 71; DB 2; Length 388; Best Local Similarity 32.5%; Pred. No. 1.5; Matches 13; Conservative 10; Mismatches 11; Indels 6; Gaps 2.	

```
Search completed: March 1, 2001, 15:52:29
Job time: 553 sec
```